

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/347,175

DATE: 08/02/2000
TIME: 12:57:04

Input Set: A:\07251056-001.txt
Output Set: N:\CRF3\08022000\I347175.raw

ENTERED

4 <110> APPLICANT: James M. Hogle
5 Harmon J. Zuccola
6 David Filman
7 Carl Elkin
9 <120> TITLE OF INVENTION: Oligomerization of Hepatitis Delta
10 Antigen
12 <130> FILE REFERENCE: 0725.1056-001
13 <140> CURRENT APPLICATION NUMBER: 09/347,175
15 <141> CURRENT FILING DATE: 1999-07-01
17 <150> PRIOR APPLICATION NUMBER: 60/091,609
18 <151> PRIOR FILING DATE: 1998-07-02
20 <160> NUMBER OF SEQ ID NOS: 35
22 <170> SOFTWARE: FastSEQ for windows Version 4.0
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 49
26 <212> TYPE: PRT
27 <213> ORGANISM: Hepatitis Delta Virus
29 <400> SEQUENCE: 1
30 Gly Arg Glu Asp Ile Leu Glu Gln Trp Val Ser Gly Arg Lys Lys Leu
31 1 5 10 15
32 Glu Glu Leu Glu Arg Asp Leu Arg Lys Leu Lys Lys Lys Ile Lys Lys
33 20 25 30
34 Leu Glu Glu Asp Asn Pro Trp Leu Gly Asn Ile Lys Gly Ile Ile Gly
35 35 40 45
36 Lys
40 <210> SEQ ID NO: 2
41 <211> LENGTH: 49
42 <212> TYPE: PRT
43 <213> ORGANISM: Hepatitis Delta Virus
45 <400> SEQUENCE: 2
46 Gly Arg Glu Glu Val Leu Glu Gln Trp Val Asn Ser Arg Lys Lys Ala
47 1 5 10 15
48 Glu Glu Leu Glu Arg Asp Leu Arg Lys Thr Lys Lys Lys Ile Lys Lys
49 20 25 30
50 Leu Glu Asp Asp Asn Pro Trp Leu Gly Asn Ile Lys Gly Ile Leu Gly
51 35 40 45
52 Lys
56 <210> SEQ ID NO: 3
57 <211> LENGTH: 49
58 <212> TYPE: PRT
59 <213> ORGANISM: Hepatitis Delta Virus
61 <400> SEQUENCE: 3
62 Gly Arg Glu Glu Val Leu Glu Gln Trp Val Ser Gly Arg Lys Lys Leu
63 1 5 10 15
64 Glu Glu Leu Glu Arg Asp Leu Arg Lys Val Lys Lys Lys Ile Lys Lys
65 20 25 30
66 Leu Glu Asp Glu His Pro Trp Leu Gly Asn Ile Lys Gly Ile Leu Gly

```

67      35      40      45
68 Lys
72 <210> SEQ ID NO: 4
73 <211> LENGTH: 49
74 <212> TYPE: PRT
75 <213> ORGANISM: Hepatitis Delta Virus
77 <400> SEQUENCE: 4
78 Gly Arg Glu Glu Val Leu Glu Gln Trp Val Ala Gly Arg Arg Lys Gln
79 1 5 10 15
80 Glu Glu Leu Glu Arg Asp Leu Arg Lys Thr Lys Lys Lys Ile Lys Lys
81 20 25 30
82 Leu Glu Glu Glu Asn Pro Trp Leu Gly Asn Ile Lys Gly Ile Leu Gly
83 35 40 45
84 Lys
88 <210> SEQ ID NO: 5
89 <211> LENGTH: 48
90 <212> TYPE: PRT
91 <213> ORGANISM: Hepatitis Delta Virus
93 <400> SEQUENCE: 5
94 Thr Arg Glu Glu Thr Leu Glu Lys Trp Ile Thr Ala Arg Lys Lys Ala
95 1 5 10 15
96 Glu Glu Leu Glu Lys Asp Leu Arg Lys Thr Arg Lys Thr Ile Lys Lys
97 20 25 30
98 Leu Glu Glu Glu Asn Pro Trp Leu Gly Asn Ile Val Gly Ile Ile Arg
99 35 40 45
102 <210> SEQ ID NO: 6
103 <211> LENGTH: 48
104 <212> TYPE: PRT
105 <213> ORGANISM: Hepatitis Delta Virus
107 <400> SEQUENCE: 6
108 Thr Arg Glu Glu Thr Leu Glu Lys Trp Ile Thr Ala Arg Lys Lys Ala
109 1 5 10 15
110 Glu Glu Leu Glu Lys Asp Leu Arg Lys Ala Arg Lys Thr Ile Lys Lys
111 20 25 30
112 Leu Glu Glu Glu Asn Pro Trp Leu Gly Asn Ile Leu Gly Ile Ile Arg
113 35 40 45
116 <210> SEQ ID NO: 7
117 <211> LENGTH: 49
118 <212> TYPE: PRT
119 <213> ORGANISM: Hepatitis Delta Virus
121 <400> SEQUENCE: 7
122 Gly Arg Glu Gln Ile Leu Glu Gln Trp Val Asp Gly Arg Lys Lys Leu
123 1 5 10 15
124 Glu Glu Leu Glu Arg Asp Leu Arg Lys Ile Lys Lys Lys Ile Lys Lys
125 20 25 30
126 Leu Glu Glu Glu Asn Pro Trp Leu Gly Asn Val Lys Gly Ile Leu Gly
127 35 40 45
128 Lys
132 <210> SEQ ID NO: 8

```

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/347,175
 DATE: 08/02/2000
 TIME: 12:57:04

Input Set : A:\07251056-001.txt
 Output Set: N:\CRF3\08022000\I347175.raw

```

133 <211> LENGTH: 49
134 <212> TYPE: PRT
135 <213> ORGANISM: Hepatitis Delta Virus
137 <400> SEQUENCE: 8
138 Gly Arg Glu Glu Ile Leu Glu Gln Trp Val Ala Gly Arg Lys Lys Leu
139 1 5 10 15
140 Glu Glu Leu Glu Arg Asp Leu Arg Lys Thr Lys Lys Lys Leu Lys Lys
141 20 25 30
142 Ile Glu Asp Glu Asn Pro Trp Leu Gly Asn Ile Lys Gly Ile Leu Gly
143 35 40 45
144 Lys
148 <210> SEQ ID NO: 9
149 <211> LENGTH: 37
150 <212> TYPE: PRT
151 <213> ORGANISM: Artificial Sequence
153 <220> FEATURE:
154 <223> OTHER INFORMATION: Residues 12-48 of delta 12-60 (Y)
156 <400> SEQUENCE: 9
157 Gly Arg Glu Asp Ile Leu Glu Gln Trp Val Ser Gly Arg Lys Lys Leu
158 1 5 10 15
159 Glu Glu Leu Glu Arg Asp Leu Arg Lys Leu Lys Lys Lys Ile Lys Lys
160 20 25 30
161 Leu Glu Glu Asp Asn
162 35
165 <210> SEQ ID NO: 10
166 <211> LENGTH: 604
167 <212> TYPE: DNA
168 <213> ORGANISM: Artificial Sequence
170 <220> FEATURE:
171 <223> OTHER INFORMATION: Synthetic Gene for Optimized Expression of HDag-S
172 in E. Coli
174 <221> NAME/KEY: CDS
175 <222> LOCATION: (7)...(591)
177 <400> SEQUENCE: 10
178 gggcat atg agc cgt agc gaa cgt cgt aaa gat cgt ggc ggc cgt gaa 48
179 Met Ser Arg Ser Glu Arg Arg Lys Asp Arg Gly Gly Arg Glu
180 1 5 10
182 gat att ctg gaa cag tgg gtg agc ggc cgt aag aag tta gag gaa ttg 96
183 Asp Ile Leu Glu Gln Trp Val Ser Gly Arg Lys Lys Leu Glu Glu Leu
184 15 20 25 30
186 gaa cgt gat ctg cgt aaa ctg aaa aag aag att aag aaa ctg gaa gaa 144
187 Glu Arg Asp Leu Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu Glu Glu
188 35 40 45
190 gat aac ccg tgg ttg ggt aat att aaa ggc att att ggc aag aaa gat 192
191 Asp Asn Pro Trp Leu Gly Asn Ile Lys Gly Ile Ile Gly Lys Lys Asp
192 50 55 60
194 aaa gat ggc gaa ggc gcg ccg ccg gcg aag aaa ctg cgt atg gat cag 240
195 Lys Asp Gly Glu Gly Ala Pro Pro Ala Lys Lys Leu Arg Met Asp Gln
196 65 70 75

```

RAW SEQUENCE LISTING

DATE: 08/02/2000

PATENT APPLICATION: US/09/347,175

TIME: 12:57:04

Input Set : A:\07251056-001.txt

Output Set: N:\CRF3\08022000\I347175.raw

```

198 atg gaa att gat gcg ggc cgg cgt aaa cgt cgg ctg cgt ggc ggc ttt 288
199 Met Glu Ile Asp Ala Gly Pro Arg Lys Arg Pro Leu Arg Gly Gly Phe
200      80      85      90
202 acc gat aag gaa cgt cag gac cat cgt cgt cgt aaa gcg ctg gaa aac 336
203 Thr Asp Lys Glu Arg Gln Asp His Arg Arg Arg Lys Ala Leu Glu Asn
204 95      100      105      110
206 aaa cgt aaa cag ctg agc agc ggc ggc aaa tct ctg agc cgt gaa gaa 384
207 Lys Arg Lys Gln Leu Ser Ser Gly Gly Lys Ser Leu Ser Arg Glu Glu
208      115      120      125
210 gaa gaa gaa ctg aaa cgt ctg acc gaa gaa gat gaa aaa cgt gaa cgt 432
211 Glu Glu Glu Leu Lys Arg Leu Thr Glu Glu Asp Glu Lys Arg Glu Arg
212      130      135      140
214 cgt att gca ggt cca tct gtt ggt ggt gtg aac ccg ctg gaa ggc ggc 480
215 Arg Ile Ala Gly Pro Ser Val Gly Gly Val Asn Pro Leu Glu Gly Gly
216      145      150      155
218 agc cgt ggt gca cgg ggc ggt ggc ttt gtg ccg tct atg caa ggt gtt 528
219 Ser Arg Gly Ala Pro Gly Gly Gly Phe Val Pro Ser Met Gln Gly Val
220      160      165      170
222 cca gaa agc ccg ttt gcg cgt acc ggc gaa ggc ctg gat att cgt ggc 576
223 Pro Glu Ser Pro Phe Ala Arg Thr Gly Glu Gly Leu Asp Ile Arg Gly
224 175      180      185      190
226 agc cag ggc ttt ccg taaaccatgg cgc 604
227 Ser Gln Gly Phe Pro
228      195
231 <210> SEQ ID NO: 11
232 <211> LENGTH: 195
233 <212> TYPE: PRT
234 <213> ORGANISM: Artificial Sequence
236 <220> FEATURE:
237 <223> OTHER INFORMATION: Amino acid sequence encoded by synthetic Gene for
238      Optimized Expression of HDAG-S in E. Coli
240 <400> SEQUENCE: 11
241 Met Ser Arg Ser Glu Arg Arg Lys Asp Arg Gly Gly Arg Glu Asp Ile
242 1      5      10      15
243 Leu Glu Gln Trp Val Ser Gly Arg Lys Lys Leu Glu Glu Leu Glu Arg
244      20      25      30
245 Asp Leu Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu Glu Glu Asp Asn
246      35      40      45
247 Pro Trp Leu Gly Asn Ile Lys Gly Ile Ile Gly Lys Lys Asp Lys Asp
248      50      55      60
249 Gly Glu Gly Ala Pro Pro Ala Lys Lys Leu Arg Met Asp Gln Met Glu
250 65      70      75      80
251 Ile Asp Ala Gly Pro Arg Lys Arg Pro Leu Arg Gly Gly Phe Thr Asp
252      85      90      95
253 Lys Glu Arg Gln Asp His Arg Arg Lys Lys Ala Leu Glu Asn Lys Arg
254      100      105      110
255 Lys Gln Leu Ser Ser Gly Gly Lys Ser Leu Ser Arg Glu Glu Glu Glu
256      115      120      125
257 Glu Leu Lys Arg Leu Thr Glu Glu Asp Glu Lys Arg Glu Arg Arg Ile

```

RAW SEQUENCE LISTING

DATE: 08/02/2000

PATENT APPLICATION: US/09/347,175

TIME: 12:57:04

Input Set : A:\07251056-001.txt

Output Set: N:\CRF3\08022000\I347175.raw

```

258      130      135      140
259 Ala Gly Pro Ser Val Gly Gly Val Asn Pro Leu Glu Gly Gly Ser Arg
260 145      150      155      160
261 Gly Ala Pro Gly Gly Gly Phe Val Pro Ser Met Gln Gly Val Pro Glu
262      165      170      175
263 Ser Pro Phe Ala Arg Thr Gly Glu Gly Leu Asp Ile Arg Gly Ser Gln
264      180      185      190
265 Gly Phe Pro
266      195
269 <210> SEQ ID NO: 12
270 <211> LENGTH: 1679
271 <212> TYPE: DNA
272 <213> ORGANISM: Hepatitis Delta Virus
274 <400> SEQUENCE: 12
275 cttgagccaa gttccgagcg aggagacgcg gggggaggat cagctcccga gaggggatgt 60
276 caccgtaaaq aqcattggaa cgtcggagaa actactccca agaagcaaaq agagggtctca 120
277 ggaagcggac gagatcccga caacgcggga gaatctctgg aaggggaaaq aggaagggtgg 180
278 aagaaaaaag ggcggggcctc ccgatccgag gggcccaacc tccagatctg gagagcactc 240
279 cggcccgaag ggttgagtag caccagagg gaggaatcca ctccgagatg agcagagaaa 300
280 tcacctccag aggaccctct cagcgaacaa gaggcgcttc gagcggtagg agtaagacca 360
281 tagcgatagg aggagatgct aggagtaggg ggagaccgaa gcgaggagga aagtaagaaa 420
282 agcaacgggg ctagccgggtg ggtgttcgcg ccccagagag gggacgagtg aggcctatcc 480
283 cgggggaattc gacttatcgt ccccatctag cgggaccccg gaccccttc gaaagtgacc 540
284 ggagggggtg ctgggaacac cggggaccag tggagccatg ggatgccctt cccgatgtct 600
285 gactccgact cccccccca aggtctgcgc aggaatggcg ggacccact ctgcagggtc 660
286 cgcgttccat cctttcttac ctgatggcgg gcctggctcc agcctcctcg ctggcgcccg 720
287 ctgggcaaca ttccgagggg accgtccctt cggtaatggc gaatgggacc cacaatatct 780
288 tctagattcc gatagagaat cgagagaaaa gtgctctcc cttagccatc cgagtggacg 840
289 tgcgtcctcc ttccgatgcc caggtcggac cgcgaggagg tggagatgcc atgccgacc 900
290 gaagaggaaa gaaggacgcg agacgcaaac ctgtgagtg aaaccgcctt tattcactgg 960
291 ggtcgacaac tctggggaga aaaggggcga tcggctggga agagtatac ccattgaaa 1020
292 ccttggttcc cctgatgtc cagccctcc cggctccgag agaaggggga ctccgggact 1080
293 ccttcgagac tggggacgaa gccgccccg ggcgctcccc tcgatccacc ttccaggggg 1140
294 ttcacacccc caaccggcg gccggtact cttctttccc ttctctctg ttctcgggtc 1200
295 aacctcctga gttctcttc ttctctctt ctgaggttct tgcctcccg cgtatgctgc 1260
296 ttctttctgt tctcagggcg cttctctct cgggtgacct gcctctctt gtccgtgaat 1320
297 cctccccctg agggcctctt cctaggtccg gagtctacct ccattctgt cgttcgggac 1380
298 ctcttcgccc ggggagcccc ctctccatcc ttatcttct ttccgagaa tcttttgatg 1440
299 ttccccagcc agggatttct gtctcaatc ttcttgagtt tcttcttct ctccggagg 1500
300 tctctctcga gttctctaa cttctttctt ccggccaccc actgctcgag gatctctct 1560
301 ctccccccgc ggttctctct cgaactcgac cggctcatct cggctagagg cggcagtcct 1620
302 cagtactctt actcttttct gtaaagagga gactgctgga ctgcgccccc gagcccgag 1679
304 <210> SEQ ID NO: 13
305 <211> LENGTH: 1683
306 <212> TYPE: DNA
307 <213> ORGANISM: Hepatitis Delta Virus
309 <400> SEQUENCE: 13
310 atgggccaag ttccgaacaa ggtccgcgg ggaggacgga tcacctcccg agaggggtaa 60
311 gtcgctaaag aqcattggaa cgtcggagat acaactccca agaagaaaaa aagagaaagc 120

```

VERIFICATION SUMMARY DATE: 08/02/2000
PATENT APPLICATION: US/09/347,175 TIME: 12:57:05

Input Set : A:\07251056-001.txt
Output Set: N:\CRF3\08022000\I347175.raw

L:736 M:361 W: Invalid Split Codon, Sequence data for SEQ ID#: 23